



PATENT

Attorney Docket 4432US

NOTICE OF EXPRESS MAILING

Express Mail Mailing Label Number: EL500245761US

Date of Deposit with USPS: June 16, 2000

Person making Deposit: Jared Turner

APPLICATION FOR LETTERS PATENT

for

PEPTIDES AND NUCLEIC ACIDS DERIVED FROM *EISENIA FOETIDA*
AND THE USE THEREOF

Inventors:

Patrick de Baetselier
Alain Beschin

Attorney:
Allen C. Turner
Registration Number: 33,041
Trask Britt
P.O. Box 2550
Salt Lake City, Utah 84111
(801) 532-1922

PEPTIDES AND NUCLEIC ACIDS DERIVED FROM *EISENIA FOETIDA* AND THE USE THEREOF

The invention relates to *Eisenia foetida* polypeptides and peptides, particularly to recombinant polypeptides useful in tumour therapy, microbial infection, inflammation or immunology.

The invention also relates to a process for preparing the above-said polypeptides and peptides.

Furthermore, the invention concerns nucleic acids coding for said polypeptides and peptides.

Tumour Necrosis Factor α (TNF- α) is a multifunctional cytokine, produced in vertebrates, mainly by activated macrophages. *In vitro*, it has several biological effects, including killing of transformed cells and antiparasitic effects. Moreover, TNF- α has been shown to have a lectin-like property for the oligosaccharide ligands chitobiose and Man(α 1,3)-Man(α 1,6)-Man (1, 2). Recently, Lucas and co-workers (3) have mapped the lectin-like domain of TNF- α and have shown that the domain exerts trypanolytic activity on salivarian trypanosomes such as *Trypanosoma brucei*. The lectin-like activity of TNF- α is functionally involved in interactions with trypanosomes and possibly also with other pathogens.

The prophenoloxidase (proPO) activating system represents an important defence mechanism in a large variety of invertebrates (4, 5). This system is based on the recognition of bacterial antigens such as lipopolysaccharide (LPS), or peptidoglycan and β -1,3-glucan, present as major components of the cell wall of yeasts and fungi (6, 7). Generally, upon the recognition of such carbohydrates proteinases cleave by limited proteolysis inactive proPO to its active state, phenoloxidase (PO). The active enzyme catalyses the o-hydroxylation of monophenols as well as the oxidation of diphenols to quinones which are subsequently polymerised non-enzymatically to melanin.

Melanin and its precursors involved in the proPO activating system have cytotoxic and antimicrobial properties and participate in a wide range of other biological activities including phagocytosis or opsonization, encapsulation or nodule formation, degranulation and wound healing (8-11).

The prophenoloxidase activating system has been detected both in protostomian and deuterostomian species. Though proPO activating system is well documented in arthropods, data in other protostomian groups are more scarce. In annelids, melanization reactions and formation of "brown bodies" or nodules have been described in polychaetes and oligochaetes (12-16). However, biochemical detection of PO activity was so far restricted to a few species with rather controversial results. While Smith and Söderhäll (17) failed to detect proPO system in the polychaete *Aphrodite aculeata* and *Arenicola marina*, Fischer (18), Valembois *et al.* (19), and Porchet-Henneré and Vernet (15) have evidenced PO activity in *Lumbricus terrestris*, *Eisenia foetida andrei* and *Nereis diversicolor* respectively. More recently, using L-DOPA as substrate, a 30 kDa protein responsible for PO activity was identified in the coelomic fluid of *Lumbricus rubellus* (20). A report showing that the oxidative activity of the coelomic fluid of earthworms towards L-DOPA *in vitro* is not affected by trypsin but completely blocked by subtilisin reflects the importance of a correct proteolytic digestion as an initial step for inactive proPO activation (19).

Since the factor which recognises microbial carbohydrates and triggers the proPO system has not yet been described in annelids (4, 5), investigations were initiated to identify such a molecule in the coelomic fluid (CF) of *E. Foetida*. Surprisingly, it is shown in this invention that a previously described 42 kDa cytolytic protein named CCF-1 (Coelomic Cytolytic Factor 1) (21) binds LPS and β -1,3 glucan and that the same protein is also responsible for the trypanolytic activity of the coelomic fluid. By combining the glucan and LPS binding capacity with the cytolytic and trypanolytic activity, the invertebrate factor resembles the vertebrate compound TNF- α and can therefore be considered as a primitive type of cytokine, which may be useful as an alternative for TNF- α . This idea is supported by the fact that an anti-

TNF monoclonal antibody (anti-TNF/TIP) crossreacts with CCF-1, whereas an anti-CCF-1 monoclonal antibody (12C9) crossreacts with TNF- α . Moreover, in *E. foetida*, CCF-1 levels are increased after LPS treatment, which resembles the TNF induction by LPS, noticed in vertebrates. Apart from the above described characteristics, it is shown that CCF-1 also participates in the proPO cascade of the coelomic fluid of *Eisenia foetida*.

Even more surprisingly, the cytolytic, trypanolytic and glucan-binding characteristics of the protein can be attributed to a small domain of 13 amino acids as depicted in SEQ.ID.NO.1. Moreover, this isolated peptide of 13 amino acids shows biological activity. The sequence of this peptide, however, is completely different from the TIP region of TNF- α , although it shares some functional characteristics.

Detailed description of the invention

It is an object of the current invention to provide new *Eisenia foetida* polypeptides and their corresponding nucleic acids which can be used in tumour therapy, microbial infection, inflammation or immunology.

It is another object of the invention to provide a nucleic acid coding for the peptide or polypeptide chains of biologically pure, active recombinant peptides which enable their preparation on large scale. When this nucleic acid encoding for the (poly)peptide is operably linked to an appropriate control sequence - such as a promoter- several host organisms, such as *E.coli*, *Bacillus sp.*, *Streptomyces sp.*, yeast, fungi, insect cells, plant cells or mammalian cells can be used for the production of the recombinant protein. Alternatively, the peptides may be produced by chemical synthesis.

A peptide or polypeptide according to this invention is characterised by the fact that it contains at least the 13 amino acids as depicted in SEQ.ID.NO.1 and preferably comprises at least 9 contiguous amino acids of said SEQ.ID.NO.1.

To the scope of the invention also relates the polypeptide of SEQ.ID.NO.3 or functional parts thereof.

According to another embodiment of the invention, the above defined peptides or polypeptides are exerting a trypanocidal or trypanolytical activity on *T. brucei* and/or *T. cruzi*, alone or preferably in combination with one of the following characteristics:

- exerting cytolytical activity
- exhibiting a β -1,3 glucan binding capacity and/or a LPS binding capacity.
- exerting opsonizing and/or hemolytic activity
- participating in the proPO cascade of *E. foetida*.

Accordingly another embodiment of the invention is the use of a peptide or polypeptide comprising at least 9 contiguous amino acids of SEQ.ID.NO.1, such as e.g. the polypeptide given by SEQ.ID.NO.3 for the manufacturing of a medicament to treat trypanosomal infection, bacterial infection or cancer. For the treatment of cancer, the peptide or polypeptide can be linked to a tumour specific antibody that directs the molecule to the tumor where the (poly)peptide can exert its cytolytical activity.

Another embodiment of the invention is the use of a peptide or polypeptide comprising at least 9 contiguous amino acids of SEQ.ID.NO.1 for the preparation of a medicament to treat a trypanosomal infection, a bacterial infection or cancer.

The invention thus relates to a DNA sequence encoding an *Eisenia foetida* protein or polypeptide or encoding an immunologically active and/or functional fragment thereof selected from the group consisting of

- (a) DNA sequences comprising a nucleotide sequence encoding a protein or peptide comprising the amino acid sequence as given in SEQ ID NO. 1 or 3;
- (b) DNA sequences comprising a nucleotide sequence as given in SEQ ID NO: 2;
- (c) DNA sequences hybridizing with the complementary strand of a DNA sequence as defined in (a) or (b) and encoding an amino acid sequence which is at least 80% identical to the amino acid

sequence encoded by the DNA sequence of (a) or (b);

- (d) DNA sequences the nucleotide sequence of which is degenerated as a result of the genetic code to a nucleotide sequence of a DNA sequence as defined in any one of (a) to (c); and
- (e) DNA sequences encoding a fragment of a protein encoded by a DNA sequence of any one of (a) to (d).

The present invention also relates to vectors, particularly plasmids, cosmids, viruses, bacteriophages and other vectors used conventionally in genetic engineering that contain a nucleic acid molecule alternatively called nucleic acid sequence or DNA sequence according to the invention. Methods which are well known to those skilled in the art can be used to construct various plasmids and vectors; see, for example, the techniques described in Sambrook, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory (1989) N.Y. and Ausubel, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. (1989). Alternatively, the nucleic acid molecules and vectors of the invention can be reconstituted into liposomes for delivery to target cells.

In a preferred embodiment the nucleic acid molecule present in the vector is linked to (a) control sequence(s) which allow the expression of the nucleic acid molecule in prokaryotic and/or eukaryotic cells.

The term "control sequence" refers to regulatory DNA sequences which are necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism. In prokaryotes, control sequences generally include promoter, ribosomal binding site, and terminators. In eukaryotes generally control sequences include promoters, terminators and, in some instances, enhancers, transactivators or transcription factors. The term "control sequence" is intended to include, at a minimum, all components the presence of which are necessary for expression, and may also include additional advantageous components.

The term "operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences. In case the control sequence is a promoter, it is obvious for a skilled person that double-stranded nucleic acid is used.

Thus, the vector of the invention is preferably an expression vector. An "expression vector" is a construct that can be used to transform a selected host cell and provides for expression of a coding sequence in the selected host. Expression vectors can for instance be cloning vectors, binary vectors or integrating vectors. Expression comprises transcription of the nucleic acid molecule preferably into a translatable mRNA. Regulatory elements ensuring expression in prokaryotic and/or eukaryotic cells are well known to those skilled in the art. In the case of eukaryotic cells they comprise normally promoters ensuring initiation of transcription and optionally poly-A signals ensuring termination of transcription and stabilization of the transcript. Additional regulatory elements may include transcriptional as well as translational enhancers. Possible regulatory elements permitting expression in prokaryotic host cells comprise, e.g., the P_L , *lac*, *trp* or *tac* promoter in *E. coli*, and examples of regulatory elements permitting expression in eukaryotic host cells are the *AOX1* or *GAL1* promoter in yeast or the CMV-, SV40-, RSV-promoter (Rous sarcoma virus), CMV-enhancer, SV40-enhancer or a globin intron in mammalian and other animal cells. In this context, suitable expression vectors are known in the art such as Okayama-Berg cDNA expression vector pcDV1 (Pharmacia), pCDM8, pRc/CMV, pcDNA1, pcDNA3 (In-vitrogen), pSPORT1 (GIBCO BRL). Advantageously, the above-described vectors of the invention comprises a selectable and/or scorable marker. Selectable marker genes are well known to those skilled in the art and comprise, for example, *npt* which confers resistance to the aminoglycosides neomycin, kanamycin and paromycin (Herrera-Estrella,

EMBO J. 2 (1983), 987-995) and hygromycin resistance (Marshall, Gene 32 (1984), 481-485).

Useful scorable markers are also known to those skilled in the art and are commercially available. Advantageously, said marker is a gene encoding luciferase (Giacomini, Pl. Sci. 116 (1996), 59-72; Scikantha, J. Bact. 178 (1996), 121), green fluorescent protein (Gerdes, FEBS Lett. 389 (1996), 44-47) or β -glucuronidase (Jefferson, EMBO J. 6 (1987), 3901-3907). This embodiment is particularly useful for simple and rapid screening of cells, tissues and organisms containing a vector of the invention.

The present invention furthermore relates to host cells comprising a vector as described above or a nucleic acid molecule according to the invention wherein the nucleic acid molecule is foreign to the host cell.

By "foreign" it is meant that the nucleic acid molecule is either heterologous with respect to the host cell, this means derived from a cell or organism with a different genomic background, or is homologous with respect to the host cell but located in a different genomic environment than the naturally occurring counterpart of said nucleic acid molecule. This means that, if the nucleic acid molecule is homologous with respect to the host cell, it is not located in its natural location in the genome of said host cell, in particular it is surrounded by different genes. In this case the nucleic acid molecule may be either under the control of its own promoter or under the control of a heterologous promoter. The vector or nucleic acid molecule according to the invention which is present in the host cell may either be integrated into the genome of the host cell or it may be maintained in some form extrachromosomally.

The host cell can be any prokaryotic or eukaryotic cell, such as bacterial, insect, fungal, plant or animal cells. Preferred fungal cells are, for example, those of the genus *Saccharomyces*, in particular those of the species *S. cerevisiae*.

Another subject of the invention is a method for the preparation of an *Eisenia foetida* protein which comprises the cultivation of host cells according to the invention which, due to the presence of a vector or a nucleic acid molecule according to the invention, are able to express such a protein, under conditions which allow expression of the protein and recovering of the so-produced protein from the culture.

The term "expression" means the production of a protein or nucleotide sequence in the cell. However, said term also includes expression of the protein in a cell-free system. It includes transcription into an RNA product, post-transcriptional modification and/or translation to a protein product or polypeptide from a DNA encoding that product, as well as possible post-translational modifications. Depending on the specific constructs and conditions used, the protein may be recovered from the cells, from the culture medium or from both. For the person skilled in the art it is well known that it is not only possible to express a native protein but also to express the protein as fusion polypeptides or to add signal sequences directing the protein to specific compartments of the host cell, e.g., ensuring secretion of the peptide into the culture medium, etc. Furthermore, such a protein and fragments thereof can be chemically synthesized and/or modified according to standard methods.

The terms "protein" and "(poly)peptide" used in this application are interchangeable. "(Poly)peptide" refers to a polymer of amino acids (amino acid sequence) and does not refer to a specific length of the molecule. Thus peptides and oligopeptides are included within the definition of polypeptide. This term does also refer to or include post-translational modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

The present invention furthermore relates to proteins encoded by the nucleic acid molecules according to the invention or produced or obtained by the above-described methods, and to functional and/or immunologically active fragments of such proteins. The proteins and polypeptides of the present invention are not necessarily translated from a designated nucleic acid sequence; the polypeptides may be generated in any manner, including for example, chemical synthesis, or expression of a recombinant expression system, or isolation from a suitable viral system. The polypeptides may include one or more analogs of amino acids, phosphorylated amino acids or unnatural amino acids. Methods of inserting analogs of amino acids into a sequence are known in the art. The polypeptides may also include one or more labels, which are known to those skilled in the art. In this context, it is also understood that the proteins according to the invention may be further modified by conventional methods known in the art. By providing the proteins according to the present invention it is also possible to determine fragments which retain biological activity, namely the mature, processed form. This allows the construction of chimeric proteins and peptides comprising an amino sequence derived from the protein of the invention, which is crucial for its binding activity and other functional amino acid sequences. The other functional amino acid sequences may be either physically linked by, e.g., chemical means to the proteins of the invention or may be fused by recombinant DNA techniques well known in the art.

The term "fragment of a sequence" or "part of a sequence" means a truncated sequence of the original sequence referred to. The truncated sequence (nucleic acid or protein sequence) can vary widely in length; the minimum size being a sequence of sufficient size to provide a sequence with at least a comparable function and/or activity of the original sequence referred to, while the maximum size is not critical. In some applications, the maximum size usually is not substantially greater than that required to provide the desired activity and/or function(s) of the original sequence. Typically, the truncated amino acid sequence will range from about 5 to about 60 amino acids in length. More typically, however, the sequence will be a maximum of about 50

amino acids in length, preferably a maximum of about 30 amino acids. It is usually desirable to select sequences of at least about 10, 12 or 15 amino acids, up to a maximum of about 20 or 25 amino acids. Specifically in this context is referred to SEQ.ID.NO. 1 comprising 13 amino acids showing the essential cytolytic, trypanolytic and glucan-binding characteristics comparable to the whole protein.

Furthermore, folding simulations and computer redesign of structural motifs of the protein of the invention can be performed using appropriate computer programs (Olszewski, *Proteins* 25 (1996), 286-299; Hoffman, *Comput. Appl. Biosci.* 11 (1995), 675-679). Computer modeling of protein folding can be used for the conformational and energetic analysis of detailed peptide and protein models (Monge, *J. Mol. Biol.* 247 (1995), 995-1012; Renouf, *Adv. Exp. Med. Biol.* 376 (1995), 37-45). In particular, the appropriate programs can be used for the identification of interactive sites of the protein according to the invention, its receptor, its ligand or other interacting proteins by computer assistant searches for complementary peptide sequences (Fassina, *Immunomethods* 5 (1994), 114-120). Further appropriate computer systems for the design of protein and peptides are described in the prior art, for example in Berry, *Biochem. Soc. Trans.* 22 (1994), 1033-1036; Wodak, *Ann. N. Y. Acad. Sci.* 501 (1987), 1-13; Pabo, *Biochemistry* 25 (1986), 5987-5991. The results obtained from the above-described computer analysis can be used for, e.g., the preparation of peptidomimetics of the protein of the invention or fragments thereof. Such pseudopeptide analogues of the natural amino acid sequence of the protein may very efficiently mimic the parent protein (Benkirane, *J. Biol. Chem.* 271 (1996), 33218-33224). For example, incorporation of easily available achiral - amino acid residues into a protein of the invention or a fragment thereof results in the substitution of amide bonds by polymethylene units of an aliphatic chain, thereby providing a convenient strategy for constructing a peptidomimetic (Banerjee, *Biopolymers* 39 (1996), 769-777). Superactive peptidomimetic analogues of small peptide hormones in other systems are described in the prior art (Zhang, *Biochem. Biophys. Res. Commun.* 224 (1996), 327-331). Appropriate peptidomimetics of the protein of the present invention can also be

identified by the synthesis of peptidomimetic combinatorial libraries through successive amide alkylation and testing the resulting compounds, e.g., for their binding and immunological properties. Methods for the generation and use of peptidomimetic combinatorial libraries are described in the prior art, for example in Ostresh, *Methods in Enzymology* 267 (1996), 220-234 and Dorner, *Bioorg. Med. Chem.* 4 (1996), 709-715.

Furthermore, a three-dimensional and/or crystallographic structure of the protein of the invention can be used for the design of peptidomimetic inhibitors of the biological activity of the protein of the invention (Rose, *Biochemistry* 35 (1996), 12933-12944; Rutenber, *Bioorg. Med. Chem.* 4 (1996), 1545-1558).

Furthermore, the present invention relates to antibodies specifically recognizing an *Eisenia foetida* protein according to the invention or parts, i.e. specific fragments or epitopes, of such a protein. The antibodies of the invention can be used to identify and isolate other *Eisenia foetida* proteins and genes in any organism. These antibodies can be monoclonal antibodies, polyclonal antibodies or synthetic antibodies as well as fragments of antibodies, such as Fab, Fv or scFv fragments etc. Monoclonal antibodies can be prepared, for example, by the techniques as originally described in Köhler and Milstein, *Nature* 256 (1975), 495, and Galfré, *Meth. Enzymol.* 73 (1981), 3, which comprise the fusion of mouse myeloma cells to spleen cells derived from immunized mammals. Furthermore, antibodies or fragments thereof to the aforementioned peptides can be obtained by using methods which are described, e.g., in Harlow and Lane "Antibodies, A Laboratory Manual", CSH Press, Cold Spring Harbor, 1988. These antibodies can be used, for example, for the immunoprecipitation and immunolocalization of proteins according to the invention as well as for the monitoring of the synthesis of such proteins, for example, in recombinant organisms, and for the identification of compounds interacting with the protein according to the invention. For example, surface plasmon resonance as employed in the BIAcore system can be used to increase the efficiency of phage antibodies selections, yielding a high increment of affinity from a single library of phage antibodies which bind to an epitope of

the protein of the invention (Schier, Human Antibodies Hybridomas 7 (1996), 97-105; Malmborg, J. Immunol. Methods 183 (1995), 7-13). In many cases, the binding phenomena of antibodies to antigens is equivalent to other ligand/anti-ligand binding.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 A and 1B: Trypanolytic activity of peptides

(A) Trypanolysis in function of the time (0.0125 mg/ml peptide)

(B) Trypanolysis (after 3 hours) in function of the concentration.

S2B is a biotinylated, irrelevant peptide.

Figure 2: Trypanolytic activity of purified natural CCF-1 (nCCF-1) versus recombinant CCF-1 (rCCF-1), tested on *T. brucei*.

Figure 3A and 3B:

(A) Inhibition of rCCF-1-mediated trypanolytic activity by anti-CCF-1 and anti-TNF/TIP monoclonal antibodies, tested on *T. brucei*.

(B) Inhibition of rCCF-1-mediated trypanolytic activity by N,N'-diacetylchitobiose, as tested on *T. brucei*.

Figure 4: Trypanolytic activity of rCCF-1 and CF on *T. cruzi*.

Figure 5: Cytolytic activity of rCCF-1, as measured on L929 cells, in presence and absence of Actinomycin D (10 µg/ml)

Figure 6: Involvement of CCF-1 in the activation of PO. Level of L-DOPA oxidation, assessed after 6 hrs of incubation, is expressed as the ratio \pm standard deviation of OD of the sample without and with proteinase inhibitor.

(A) entire coelomic fluid

- (B) CCF-1-depleted coelomic fluid by pre-incubation on anti-CCF-1 immunoaffinity column
- (C) CCF-1 depleted coelomic fluid supplemented with 0.5 µg/ml rCCF-1
- (D) as (C), but with 1 µg/ml rCCF-1
- (E) as (C), but with 2 µg/ml rCCF-1
- (F) CCF-1 depleted coelomic fluid supplemented with 2 µg/ml bovine serum albumin

Laminarin or LPS were given at 2 µg/ml or 1 µg/ml respectively.

EXAMPLES

1. Purification and partial sequencing of CCF-1

CCF-1 was prepared and purified as described earlier (21). Analysis on 2 dimensional polyacryl amide gel electrophoresis (PAGE) confirmed the homogeneity of the CCF-1 preparation since a single spot was detected. A bulk preparation of immunoaffinity purified CCF-1 was separated on a preparative SDS-PAGE, blotted on PVDF problott and stained with amido black. 1/10th of the amount of the immobilised protein was N-terminal sequenced for 30 cycles. The N-terminal amino acid sequence of 30 residues is shown in Table 1. The remaining part of the CCF-1 containing PVDF band was used for internal sequence determination. To this end, tryptic digestion was performed according to the method of Fernandez *et al.* (22). After cleavage, the released peptides were separated on a reverse phase column (Vydac C4, 2.1 x 250 mm) and eluted with a linear gradient (0% to 70%) of acetonitrile in 0.1% trifluoroacetic acid. The column outlet was directly connected to a 1000 S diode assay detector (Applied Biosystems) and the most prominent peaks were used for amino acid sequencing. Purified peptides were sequenced using a pulsed-liquid model 477A sequenator (Applied Biosystems). The sequences of some of the peptides are shown in Table 1. From a number of peptide sequences degenerate PCR primers were deduced to identify CCF-1 cDNA from a cDNA library of *E. foetida*.

2. Trypanolytic activity of CCF-1

Coelomic fluid collection

Coelomic fluid from adults *Eisenia foetida foetida* earthworms was obtained by puncturing the coelomic cavity with a glass micropipette in presence of protease inhibitor (Pefabloc, Boehringer, 10 mM). The pooled suspension was centrifuged (100 x g, 10 min), and the cell-free coelomic fluid after re-centrifugation (3000 x g, 10 min) was stored at -70°C until used.

Monoclonal antibodies (mAb's)

Generation of anti-CCF-1 monoclonal antibodies (12C9) was previously reported (21, 24). Isotype matched monoclonal antibody 15D3 (IgG1) against *Bandeiraea simplicifolia* B4 isolectin was prepared using standard procedures and used as control in trypanolysis assays.

Parasites

Pleiomorphic and monomorphic *Trypanosoma brucei* AnTat 1.1E clone were provided by Dr N. Van Meirvenne (Institute of Tropical Medicine, Antwerp, Belgium).

Trypanolytic assay

Purified parasites were resuspended at 4×10^6 /ml PSG (PBS, consisting of 2.13 g NaCl, 8.45 g Na_2HPO_4 and 0.43 g NaH_2PO_4 per liter, supplemented with 1 % glucose). 100 μl of suspension were mixed with 100 μl of different concentrations of CF, CCF-1 or $\text{TNF-}\alpha$ in 96-well culture plate. Quantification of the trypanolysis was based on the determination of parasite viability using Ethidium homodimer (EthD-1, Molecular Probes). EthD-1 is excluded by intact plasma membrane of live parasites. However, it enters the cells with damaged membrane and undergoes enhancement of fluorescence on binding nucleic acids, thereby producing a bright red fluorescence in dead cells. Briefly, after 4.5 hrs incubation at 30°C, 25 μl of 4.5 μM EthD-1 was added to

the wells. The mean fluorescence intensity of triplicate cultures (\pm SD) was recorded 30 min later in a cytofluorimeter (Cytofluor II, PerSeptive Biosystems) using excitation / emission filters at 530 ± 25 / 645 ± 40 nm. A set of control was included in each experiments : (A) a parasite-free control to account for possible background fluorescence (B) a control of 100% lysis, prepared by treating the parasites with 10 μ l 2 % saponin for about 10 min before adding EthD-1 (C) a control in which the trypanolytic component was replaced by PSG was considered as 0 % lysis. (B) and (C) were used to build a standard curve and the % lysis in test samples was estimated by linear regression. In some assays parasite lysis was assessed by light microscopy counting the remaining parasites. Background lysis, i.e. lysis of trypanosomes within 5 hrs incubation at 30°C in absence of lytic molecules, never exceeded 5-10%.

For inhibition experiments CCF-1 or TNF- α were pre-incubated 1hr at 30°C with 10 μ g/ml of antibodies (anti-CCF1, anti-TIP or irrelevant mAb) or sugars (chitobiose, cellobiose, Sigma). All experiments were repeated at least 3 times.

Lysis of African trypanosomes

CF of *E. foetida* exerts a trypanolytic activity on *T. brucei* parasites not being due to proteolysis because all experiments were performed in the presence of a non-toxic serine proteinase inhibitor. Immunoaffinity purified CCF-1 exerts also a trypanolytic activity, 1000-times higher as compared to the total CF indicating an efficient enrichment. The trypanolytic activity of total CF as well as of CCF-1 is completely inhibited by the neutralising antibody mAb 12C9 (21) indicating that CCF-1 accounts for the total trypanolytic activity of CF (Table 2).

Since the trypanolytic activity of TNF- α is mediated by a lectin-like domain, it was tested whether CCF-1 exhibits similar features. N,N'-diacetylchitobiose, a potent inhibitor of the trypanolytic activity of TNF- α was found to inhibit also the trypanolytic activity of CCF-1. In contrast, cellobiose, that does not influence the trypanolytic activity of TNF- α , does not influence the trypanolytic

activity of CCF-1 either. These results indicate that the CCF-1/trypanosome interaction involves a lectin-like activity. The lectin-like activity of TNF- α is mediated by a distinct domain of the molecule encompassing the amino acid sequence T¹⁰⁴PEGAE¹⁰⁹, designated as the TIP region of TNF- α (3). Since anti-TIP antibodies strongly inhibit the trypanolytic but not the cytolytic activity of TNF- α , the influence of such antibodies (polyclonal and monoclonal) was tested on the trypanolytic activity of CCF-1. According to these results, anti-TIP antibodies inhibit potently this activity (Table 3). Another monoclonal anti-TNF- α antibody (1F3F3) that neutralises potently the cytolytic (3) but weakly the trypanolytic activity of TNF- α does not interfere with CCF-1-mediated trypanolysis. These results suggest that CCF-1 and TNF- α share a similar region that mediates with interaction trypanosomes. This possibility was further on substantiated by following observations: (1) the neutralising anti-CCF-1 mAb 12C9 inhibits the trypanocidal activity of TNF- α (Table 3), (2) monoclonal anti-TIP antibodies cross-react with CCF-1 in western blot and conversely the anti-CCF-1 mAb 12C9 binds on immobilised TNF- α , (3) both CCF-1 and TNF- α bind to immobilized chitobiose.

Lysis of American trypanosomes

It has been demonstrated that TNF- α exerts also trypanolytic activity on the American trypanosome *T. cruzi*. This trypanolytic activity can be inhibited by N,N'-diacetylchitobiose and by anti-TNF/TIP antibodies but not by the anti-TNF- α mAb 1F3F3 that neutralises the cytolytic effects of TNF- α .

In view of these results, we have tested whether CF is trypanolytic for *T. cruzi*. The results are summarised in Table 4. CF contains components that lyse *T. cruzi* and this lytic activity is substantially decreased by the neutralising anti-CCF-1 mAb 12C9 and by N,N'-diacetylchitobiose, but not by cellobiose. Hence the results indicate that CCF-1 is the major *T. cruzi* trypanolytic molecule of total CF and that this activity is again mediated via a lectin-like interaction similar to that one of TNF- α . Interestingly the *T. cruzi* trypanolytic activity of both TNF- α and CCF-1 could only be recorded on the trypomastigote but not the epimastigote forms of the parasite indicating that

the susceptibility towards the lytic activity of TNF- α and CCF-1 is developmentally regulated.

3. Cytolytic activity of CCF-1

To test whether the cytolytic activity of CCF-1 is mediated by a similar or a different domain as that one utilised for the trypanolytic activity, CCF-1 was preincubated with N,N'-diacetylchitobiose or anti-TNF/TIP antibodies and tested in the L-929 lysis assay (21). The results are shown in Table 5. These results clearly demonstrate that the interaction of CCF-1 with L-929 cells is completely inhibited by treatment with N,N'-diacetylchitobiose and anti-TNF/TIP antibodies. This is in sharp contrast to the cytolytic activity of TNF- α that is not influenced by anti-TNF/TIP antibodies nor by N,N'-diacetylchitobiose. Hence, CCF-1 utilises its lectin-like domain to interact with trypanosomes and mammalian (L-929) cells and this interaction leads to cellular lysis.

4. Identification of the trypanolytic domain of CCF-1

The separated peptides obtained from HPLC purification of the tryptic digest of CCF-1 were analysed for reactivity with the anti-CCF-1 mAb via a dot spot assay. One peptide scored weakly positive with the 12C9 mAb and this peptide was sequenced. The obtained sequence is shown in SEQ.ID.NO.1. This peptide (termed CCF-1/TIP), which is clearly different from the TNF/TIP, was tested in trypanolysis assay and was found to be trypanolytic in a time- and dose-dependent way (Fig. 1). Furthermore, the CCF-1/TIP was reproducibly more trypanolytic than the TNF/TIP peptide, while irrelevant peptides were not or marginally trypanolytic (Fig. 1).

Profile analysis of the CCF-1/TIP sequence against the Swiss-prot database revealed that the CCF-1/TIF peptide contains a pattern that is unique for glycosyl hydrolases.

5. Cloning and expression of CCF-1

Molecular cloning

- *RNA extraction*

Earthworms were taken from compost and put on filter paper soaked with PBS for four days in order to clean their gut. After two days, Penicillin/Streptomycin (100U/100 µg/ml) were added to decrease the amount of micro-organisms in the worms and the worms were treated with LPS, a known inducer of CCF-1 synthesis. Three worms were frozen in liquid nitrogen, ground to powder and total RNA was prepared as described by Chomczynski and Sacchi (23). Poly A⁺ RNA was isolated from the total RNA using the polyAtract system (Promega). Biotinylated-Oligo(dT) was added to the total RNA and allowed to bind the poly A⁺ RNA. Using streptavidin-paramagnetic beads, the oligo(dT)-mRNA hybrids were captured and finally the mRNA was eluted. The mRNA isolated yielded 8 µg mRNA in 250 µl. After analysis on agarose gel, a smear of mRNA was observed and the rRNA was almost completely removed.

- *cDNA library construction*

cDNA was synthesized using a cDNA synthesis module (Amersham).

To make a cDNA library, the lambda zap II vector of Stratagene was used. Before inserting the cDNA in the lambda vector, *EcoR* I adaptors were ligated onto the ends of the cDNA. The adapted cDNA was size fractionated, phosphorylated and ligated in the lambda arms. The recombinant lambda DNA was packaged *in vitro* with Gigapack II Gold Packaging extracts (Stratagene). *E. coli* (strain XL1-Blue MRF', Stratagene) was infected with the phage suspension for amplification and determination of the phage titer.

- *Isolation of CCF-1 cDNA*

Degenerate primers were deduced from a number of native earthworm CCF-1 peptide sequences in order to identify the CCF-1 cDNA from the *E. foetida* library. Combination of the following primers (5'TIACIGAITGGGAICAA/GTAT/CATIGTITGGCA3' and 5'AAIGTITGIAAITTG/ATCICCG/ATAG/ATTCCA3') resulted in a specific PCR

fragment. A DIG labelled PCR product (DIG labelling mix, Boehringer, Mannheim) was subsequently used as a CCF-1 specific probe in plaque hybridisation. Positive plaques were isolated and the presence of CCF-1 cDNA was confirmed with PCR. In order to obtain pBluescript phagemids *in vivo* excision was performed according to Stratagene's recommendations.

- *Sequence analysis of the CCF-1 cDNA*

Since the 2500 bp *EcoR* I insert was too large to be sequenced in one run and since no information was available on the DNA sequence of the insert, deletions were made of the clone. With a Erase-a-base system (Promega) a series of deletions of the *EcoR* I insert of the pBluescript vector were made in which the insert was each time 250 bp smaller. Using these deletions it was possible to sequence the 2500 bp *EcoR* I insert of the pBluescript vector.

- *Analysis of CCF-1 cDNA*

The sequencing revealed that the *EcoR* I insert was 2682 bp in length. In the insert two open reading frames (ORF) were recognised. One ORF showed a strong homology with cDNA of the Myosin essential light chain of the earthworm *Lumbricus terrestris*. In the second ORF, all identified sequences, including the CCF-1/TIP sequence, of the natural CCF-1 of *E. foetida* were present. These results indicate that the second ORF is the complete cDNA of CCF-1. Further analysis of the CCF-1 cDNA showed that the full length cDNA was isolated with a length of 1155 bp, starting with an ATG initiation codon and ending with a TGA stop codon (SEQ.ID.NO. 2). The deduced amino acid sequence contains an eukaryotic signal sequence of 17 amino acids, indicating that the mature CCF-1 starts at amino acid 18. The deduced amino acid sequence contains 3 cysteins, possibly forming one sulphur bridge. There are no N-glycosylation sites present, excluding N-glycosylation of the protein. Since the molecular weight of the natural CCF-1 and of the deduced amino acid sequence are similar, there are probably no O-glycosylations present.

GAC AAC AAT GCC ATG CAG GTC GAT TAC ATC CGA GTC TAC AAG CGC AAC
1152
Asp Asn Asn Ala Met Gln Val Asp Tyr Ile Arg Val Tyr Lys Arg Asn
355 360 365

(2) INFORMATION FOR SEQ ID NO: 3:

(A) LENGTH: 384 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Met -17	Arg	Trp -15	Thr	Leu	Val	Val	Leu -10	Cys	Leu	Leu	Phe	Gly -5	Glu	Gly	Phe
Ala	Phe 1	Thr	Asp	Trp	Asp 5	Gln	Tyr	His	Ile	Val 10	Trp	Gln	Asp	Glu	Phe 15
Asp	Tyr	Phe	Asp	Gly 20	Ala	Lys	Trp	Gln	His 25	Glu	Val	Thr	Ala	Thr 30	Gly
Gly	Gly	Asn 35	Ser	Glu	Phe	Gln	Leu	Tyr 40	Thr	Gln	Asp	Gly	Ala 45	Asn	Ser
Phe	Val	Arg 50	Asp	Gly	Lys	Leu	Phe 55	Ile	Lys	Pro	Thr	Leu 60	Leu	Ala	Asp
Asn	Ile 65	Asn	Pro	Gln	Thr	Gly 70	Ala	Pro	Phe	Gly	Thr 75	Asp	Phe	Met	Tyr
Asn 80	Gly	Val	Leu	Asp	Val 85	Trp	Ala	Met	Tyr	Gly 90	Ala	Cys	Thr	Asn	Thr 95
Asp	Asn	Asn	Gly	Cys 100	Tyr	Arg	Thr	Gly	Ala 105	Ala	Gly	Asp	Ile	Pro 110	Pro
Ala	Met	Ser	Ala 115	Arg	Val	Arg	Thr	Phe 120	Gln	Lys	Tyr	Ser	Phe 125	Thr	His
Gly	Arg	Val 130	Val	Val	His	Ala	Lys 135	Met	Pro	Val	Gly	Asp 140	Trp	Leu	Trp
Pro	Ala 145	Ile	Trp	Met	Leu	Pro 150	Glu	Asp	Trp	Val	Tyr 155	Gly	Gly	Trp	Pro
Arg 160	Ser	Gly	Glu	Ile	Asp 165	Ile	Ile	Glu	Thr	Ile 170	Gly	Asn	Arg	Asp	Phe 175
Lys	Asn	Thr	Gly	Gly 180	Glu	Phe	Leu	Gly	Ile 185	Gln	Lys	Met	Gly	Ser 190	Thr
Met	His	Trp	Gly 195	Pro	Gly	Trp	Asp	Asp 200	Asn	Arg	Tyr	Trp	Leu 205	Thr	Ser

Leu Pro Lys His Ser Asp Asp Trp Asn Tyr Gly Asp Asn Phe His Thr
 210 215 220
 Phe Trp Phe Asp Trp Ser Pro Asn Gly Leu Arg Phe Phe Val Asp Asp
 225 230 235
 Glu Asn Gln Ala Leu Leu Asp Val Pro Tyr Pro Leu Ile Asp Ala Asn
 240 245 250 255
 Pro Trp Trp Val Asp Phe Trp Glu Trp Gly Lys Pro Trp Leu Pro Gln
 260 265 270
 Tyr Glu Asn Asp Asn Pro Trp Ala Gly Gly Thr Asn Leu Ala Pro Phe
 275 280 285
 Asp Gln Asn Phe His Phe Ile Leu Asn Val Ala Val Gly Gly Thr Asn
 290 295 300
 Gly Phe Ile Pro Asp Gly Cys Ile Asn Arg Gly Gly Asp Pro Ala Leu
 305 310 315
 Gln Lys Pro Trp Ser Asn Gly Asp Trp Tyr Asn Asp Ala Met Arg Lys
 320 325 330 335
 Phe Phe Asp Ala Arg Gly Asn Trp Lys Trp Thr Trp Asp Asp Glu Gly
 340 345 350
 Asp Asn Asn Ala Met Gln Val Asp Tyr Ile Arg Val Tyr Lys Arg Asn
 355 360 365